

# ON TRACTABILITY OF APPROXIMATION FOR A SPECIAL SPACE OF FUNCTIONS

M. HEGLAND AND G. W. WASILKOWSKI

ABSTRACT. We consider approximation problems for a special space of  $d$  variate functions. We show that the problems have small number of active variables, as it has been postulated in the past using *concentration of measure* arguments. We also show that, depending on the norm for measuring the error, the problems are strongly polynomially or quasi-polynomially tractable even in the model of computation where functional evaluations have the cost exponential in the number of active variables.

## 1. INTRODUCTION

This paper is inspired by [4], where an importance of a special class of multivariate functions was advocated, and by recent results on tractability of problems dealing with infinite-variate functions, see [1, 2, 5, 6, 9, 10, 13, 15, 16, 17], where the cost of an algorithm depends on the number of active variables that it uses.

The selection of functions in [4] was based on a particular choice of the metric used in the space of the variables  $x_i$  of the functions and on the smoothness of the functions. Here we consider the case where the  $x_i$  denote features of some objects. Adding new features will increase the distance in general, and this increase can grow substantially with the dimension. For example, if  $x_i \in [0, 1]$  for  $i = 1, \dots, d$  then the average squared Euclidean distance of two points grows proportional to the dimension  $d$ :

$$\int_{[0,1]^d} \int_{[0,1]^d} \sum_{i=1}^d (x_i - y_i)^2 \, d\mathbf{x} \, d\mathbf{y} = O(d).$$

This unbounded growth shows that Euclidean distance cannot approximate any distance function between two objects for large  $d$ . This is why it was suggested in [4] to use a scaled Euclidean distance to characterize the dissimilarity of two objects based on features  $x_1, \dots, x_d$ :

$$\text{dist}(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{1}{d} \sum_{i=1}^d (x_i - y_i)^2}.$$

The continuity of functions considered in [4] was Lipschitz-continuity based on the scaled Euclidean distance. For differentiable functions, this leads to conditions of bounded

$$\sum_{i=1}^d d \cdot \left( \frac{\partial f}{\partial x_i} \right)^2 \leq L_1$$

where  $L_1$  is the Lipschitz constant of  $f$  with respect to the scaled Euclidean distance. A model example is the mean function

$$f(\mathbf{x}) = \frac{1}{d} \sum_{i=1}^d x_i.$$

This function has a Lipschitz constant  $L_1 = 1$ . Consequently the gradient satisfies

$$\|\nabla f\|_{L_2} \leq \frac{1}{d^{1/2}}.$$

It follows that  $f$  is approximated with an  $O(d^{-1/2})$  error by the constant 0.5, i.e., the values of  $f$  are concentrated around 0.5. This concentration phenomenon for general Lipschitz-continuous functions was established by Lévy in [8].

Higher order approximations can be derived in the case when higher order Lipschitz constants are finite, i.e., where for some  $m > 0$  one has

$$\sum_{i_1 \leq \dots \leq i_m} d^m \cdot \left( \frac{\partial^m f}{\partial x_{i_1} \dots \partial x_{i_m}} \right)^2 \leq L_m^2.$$

Using the example of the mean function one has

$$\left( \frac{1}{d} \sum_{i=1}^d x_i - \frac{1}{2} \right)^2 = O(1/d).$$

From this one gets the first order (additive function) approximation

$$\frac{2}{d^2} \sum_{i < j} x_i x_j = \frac{1}{d} \sum_{i=1}^d (1 - x_i/d) x_i - 1/4 + O(1/d).$$

A similar approximation is obtained for the average squared distance  $\frac{2}{d(d-1)} \sum_{i < j} (x_i - x_j)^2$ . Both these functions do satisfy a higher order Lipschitz condition with respect to the scaled norm introduced earlier.

Classes of such functions and the particular scaling by  $1/d^m$ , where  $m$  is equal to the number of involved variables, are related to the weighted reproducing kernel Hilbert space  $\mathcal{H}_d$  of multivariate functions on  $[0, 1]^d$  with the reproducing kernel given by

$$\mathcal{K}(\mathbf{x}, \mathbf{y}) = 1 + \sum_{\mathbf{u} \neq \emptyset} d^{-|\mathbf{u}|} \cdot \prod_{j \in \mathbf{u}} \min(x_j, y_j).$$

Here the sum is over all subsets  $\mathbf{u}$  of  $\{1, \dots, d\}$ . This is why we consider such spaces in the current paper. It is well known, see, e.g., [7], that functions from that space have an ANOVA-like representation of the form

$$f(\mathbf{x}) = f_\emptyset + \sum_{\mathbf{u} \neq \emptyset} f_{\mathbf{u}}(\mathbf{x}),$$

where each component  $f_{\mathbf{u}}$  depends on, exactly, the variables listed in  $\mathbf{u}$ . Hence,  $\mathbf{u}$  is the list of active variables in  $f_{\mathbf{u}}$  and the scaling parameter  $m$  is equal to  $|\mathbf{u}|$ . The

corresponding norm of  $f$  is given by

$$\|f\|_{\mathcal{H}_d}^2 = |f_\emptyset|^2 + \sum_{\mathbf{u} \neq \emptyset} d^{|\mathbf{u}|} \cdot \left\| \frac{\partial^{|\mathbf{u}|} f_{\mathbf{u}}}{\prod_{j \in \mathbf{u}} \partial x_j} \right\|_{L_2}^2.$$

As already mentioned, it was also postulated in [4] that functions of this form are well approximated by sums of those components  $f_{\mathbf{u}}$  that depend on small numbers of variables, i.e., with  $\mathbf{u}$  of small cardinality, or just by a constant function. We show in a more quantified way, that this is true for approximation problems with errors measured in a norm of another Hilbert space  $\mathcal{G}_d$  that also has a tensor product form. That is, we show that to approximate  $f$  with an error not exceeding  $\varepsilon \cdot \|f\|_{\mathcal{H}_d}$ , it is enough to consider only those terms  $f_{\mathbf{u}}$  that depend on at most  $|\mathbf{u}| \leq m(\varepsilon, d)$  variables, where  $m(\varepsilon, d)$  grows with  $1/\varepsilon$  very slowly and/or decreases to zero when  $d$  tends to infinity. More precisely, for general tensor product spaces (including the  $L_2$  space), we have

$$m(\varepsilon, d) \leq \min \left( d, \frac{c \cdot \ln(1/\varepsilon)}{\ln(\ln(1/\varepsilon))} \right).$$

for a known constant  $c > 0$  that does not depend on  $\varepsilon$  and  $d$ . For instance, for any  $d \in \mathbb{N}_+$  and the error demand  $\varepsilon = 10^{-q}$ , we have

$$m(10^{-2}, d) \leq 5, \quad m(10^{-4}, d) \leq 8, \quad \text{and} \quad m(10^{-8}, d) \leq 14.$$

Suppose next that spaces  $\mathcal{H}_d$  and  $\mathcal{G}_d$  satisfy the following assumption: there exists  $C < \infty$  such that

$$(1) \quad \|f\|_{\mathcal{G}_d}^2 \leq C \cdot \sum_{\mathbf{u}} \|f_{\mathbf{u}}\|_{\mathcal{G}_d}^2 \quad \text{for all} \quad f = \sum_{\mathbf{u}} f_{\mathbf{u}} \in \mathcal{H}_d.$$

Then  $m(\varepsilon, d)$  has even a smaller upper bound

$$m(\varepsilon, d) \leq \min \left( d, \frac{2 \cdot \ln(1/\varepsilon)}{\ln(d/c)} \right).$$

Hence for a fixed error demand  $\varepsilon$ ,  $m(\varepsilon, d) = O(1/\ln(d))$  as  $d \rightarrow \infty$ .

Actually, we prove these results for reproducing kernels of the form

$$\mathcal{K}(\mathbf{x}, \mathbf{y}) = 1 + \sum_{\mathbf{u} \neq \emptyset} d^{-|\mathbf{u}|} \cdot \prod_{j \in \mathbf{u}} K(x_j, y_j)$$

for a general class of univariate kernels  $K : D \times D \rightarrow \mathbb{R}$  including of course  $K(x, y) = \min(x, y)$  and  $D = [0, 1]$ .

We also study the tractability of approximation problems for algorithms that can use arbitrary linear functional evaluations. However, as it has been done in the recent study of infinite-variate problems, we assume that the cost of each such evaluation depends on the number  $k$  of active variables and is given by  $\$(k)$ . Under the general tensor product assumption, approximation is quasi-polynomially tractable, and it is strongly polynomially tractable if (1) is satisfied. These results hold even when the cost function  $\$$  is exponential. We also find a sharp upper bound on the exponent of strong tractability. Approximation is weakly tractable even when  $\$$  is doubly exponential.

## 2. BASIC DEFINITIONS

**2.1. Space of  $d$ -Variate Functions.** Let  $D \subseteq \mathbb{R}$  be a Borel measurable set and let  $H = H(K)$  be a reproducing kernel Hilbert space (RKH space for short) of functions  $f : D \rightarrow \mathbb{R}$  whose kernel is denoted by  $K$ .

We assume that

$$1 \notin H,$$

where 1 denotes the constant function  $f(x) = 1$  for all  $x$ .

In what follows we write  $[1..d]$  to denote the set of positive integers not exceeding  $d$ ,

$$[1..d] := \{n \in \mathbb{N}_+ : n \leq d\}$$

and use  $\mathbf{u}, \mathbf{v}$  to denote subsets of  $[1..d]$ . Consider now the weights

$$(2) \quad \gamma_{d,\mathbf{u}} := d^{-|\mathbf{u}|} \quad \text{for } \mathbf{u} \subseteq [1..d].$$

Clearly  $\gamma_{d,\emptyset} = 1$ .

The weighted space of  $d$ -variate functions  $f : D^d \rightarrow \mathbb{R}$  under the consideration is the RKH space  $\mathcal{H}_d$  whose kernel is given by

$$\mathcal{K}_d(\mathbf{x}, \mathbf{y}) := \sum_{\mathbf{u} \subseteq [1..d]} \gamma_{d,\mathbf{u}} \cdot K_{\mathbf{u}}(\mathbf{x}, \mathbf{y}) \quad \text{and} \quad K_{\mathbf{u}}(\mathbf{x}, \mathbf{y}) = \prod_{j \in \mathbf{u}} K(x_j, y_j)$$

with the convention that  $K_{\emptyset} \equiv 1$ .

For each  $\mathbf{u}$ , by  $H_{\mathbf{u}}$  we denote the RKH space whose kernel is equal to  $K_{\mathbf{u}}$ . Clearly  $H_{\emptyset} = \text{span}\{1\}$  and  $H_{\mathbf{u}} \simeq H^{\otimes |\mathbf{u}|}$  for  $\mathbf{u} \neq \emptyset$ . It is well known that the spaces  $H_{\mathbf{u}}$ , as subspaces of  $\mathcal{H}_d$ , are mutually orthogonal and any  $f \in \mathcal{H}_d$  has the unique representation

$$f(\mathbf{x}) = \sum_{\mathbf{u} \subseteq [1..d]} f_{\mathbf{u}}(\mathbf{x}) \quad \text{with } f_{\mathbf{u}} \in H_{\mathbf{u}}$$

and

$$\|f\|_{\mathcal{H}_d}^2 = \sum_{\mathbf{u} \subseteq [1..d]} \|f_{\mathbf{u}}\|_{H_{\mathbf{u}}}^2 = \sum_{\mathbf{u} \subseteq [1..d]} \gamma_{d,\mathbf{u}}^{-1} \cdot \|f_{\mathbf{u}}\|_{H_{\mathbf{u}}}^2.$$

This representation is similar to the *ANOVA* decomposition since each term  $f_{\mathbf{u}}$  depends only on the variables listed in  $\mathbf{u}$ . The space considered in [4] and mentioned in the Introduction is related to space  $\mathcal{H}_d$  with the classical Wiener kernel discussed in the following example.

**Example.** Consider

$$D = [0, 1] \quad \text{and} \quad K(x, y) = \min(x, y).$$

Then  $H$  is the space of functions  $f : [0, 1] \rightarrow \mathbb{R}$  that vanish at zero, are absolutely continuous, and have  $f' \in L_2([0, 1])$ . The norm in  $H$  is given by

$$\|f\|_H^2 = \int_0^1 |f'(x)|^2 dx.$$

For  $\mathbf{u} \neq \emptyset$ ,  $H_{\mathbf{u}}$  consists of functions that depend only on the variables  $x_j$  with  $j \in \mathbf{u}$ , are zero if at least one of those variables is zero, have the mixed first-order partial

derivatives bounded in the  $L_2$  norm, and

$$\|f\|_{\mathcal{H}_d}^2 = |f(\mathbf{0})|^2 + \sum_{\mathbf{u} \neq \emptyset} d^{|\mathbf{u}|} \int_{D^d} \left| \prod_{j \in \mathbf{u}} \frac{\partial}{\partial x_j} f([\mathbf{x}; \mathbf{u}]) \right|^2 d\mathbf{x} \quad \text{for } f \in \mathcal{H}_d,$$

where  $[\mathbf{x}; \mathbf{u}]$  is given by

$$[\mathbf{x}; \mathbf{u}] = [y_1, \dots, y_d] \quad \text{with} \quad y_j := \begin{cases} x_j & \text{if } j \in \mathbf{u}, \\ 0 & \text{otherwise.} \end{cases}$$

**2.2. Function Approximation Problems.** For every  $d \geq 1$ , let  $\mathcal{G}_d$  be a separable Hilbert space of functions on  $D^d$  such that  $\mathcal{H}_d$  is continuously embedded in it. We denote the corresponding embedding operator by  $\mathcal{S}_d$ , i.e.,

$$\mathcal{S}_d : \mathcal{H}_d \rightarrow \mathcal{G}_d \quad \text{and} \quad \mathcal{S}_d(f) = f.$$

We assume that  $\mathcal{S}_d$  and  $\mathcal{G}_d$  have tensor product forms, i.e., for every  $\mathbf{u}$  and every  $f(\mathbf{x}) = \prod_{j \in \mathbf{u}} f_j(x_j)$  with  $f_j \in H$ , we have

$$(3) \quad \|f\|_{\mathcal{G}_d} = \prod_{j \in \mathbf{u}} \|f_j\|_{\mathcal{G}_1}.$$

For simplicity of presentation we also assume that

$$\|1\|_{\mathcal{G}_1} = 1 \quad \text{so that} \quad \|1\|_{\mathcal{G}_d} = 1.$$

The continuity of  $\mathcal{S}_d$  is equivalent to continuity of  $\mathcal{S}_1$ . Indeed, let

$$(4) \quad C_0 := \sup_{\|f\|_H \leq 1} \|f\|_{\mathcal{G}_1} < \infty.$$

Then for every  $\mathbf{u}$  we have

$$\sup_{\|f\|_{H_{\mathbf{u}}} \leq 1} \|f\|_{\mathcal{G}_d} = C_0^{|\mathbf{u}|}$$

and

$$\|\mathcal{S}_d\|^2 \leq \sum_{\mathbf{u} \subseteq [1..d]} \gamma_{d,\mathbf{u}} \cdot C_0^{2|\mathbf{u}|} = \sum_{k=0}^d \binom{d}{k} \cdot d^{-k} \cdot C_0^{2 \cdot k} = \left(1 + \frac{C_0^2}{d}\right)^d,$$

since

$$\|f\|_{\mathcal{G}_d}^2 \leq \left( \sum_{\mathbf{u} \subseteq [1..d]} C_0^{|\mathbf{u}|} \cdot \|f_{\mathbf{u}}\|_{H_{\mathbf{u}}} \right)^2 \leq \sum_{\mathbf{u} \subseteq [1..d]} \gamma_{d,\mathbf{u}} \cdot C_0^{2 \cdot |\mathbf{u}|} \cdot \|f\|_{\mathcal{H}_d}^2.$$

Clearly

$$1 \leq \|\mathcal{S}_d\| \leq e^{C_0^2/2} \quad \text{for every } d$$

which means that the corresponding approximation problem is properly scaled.

Note also that the condition (1) holds if

$$(5) \quad \langle 1, f \rangle_{\mathcal{G}_1} = 0 \quad \text{for all } f \in H.$$

Actually, under (5) we have

$$\|f\|_{\mathcal{G}_d}^2 = \sum_{\mathbf{u} \subseteq [1..d]} \|f_{\mathbf{u}}\|_{\mathcal{G}_d}^2 \quad \text{for all } f \in \mathcal{H}_d.$$

Then we can get a better estimate of the norm of  $\mathcal{S}_d$ :

$$\|f\|_{\mathcal{G}_d}^2 \leq \sum_{\mathbf{u}} d^{|\mathbf{u}|} \cdot \|f_{\mathbf{u}}\|_{H_{\mathbf{u}}}^2 \cdot C_0^{2 \cdot |\mathbf{u}|} \cdot d^{-|\mathbf{u}|} \leq \|f\|_{\mathcal{H}_d}^2 \cdot \max_{k \leq d} C_0^{2 \cdot k} \cdot d^{-k}.$$

Since the estimation above is sharp, we conclude that

$$\|\mathcal{S}_d\| = \max_{k \leq d} C_0^k \cdot d^{-k/2}$$

The class of such approximation problems contains the following *weighted- $L_2$*  approximation.

**2.2.1. Weighted  $L_2$  Approximation.** Let  $\rho$  be a given probability density function (p.d.f. for short) on  $D$ . Without loss of generality, suppose that  $\rho$  is positive (a.e.) on  $D$ . Then the  $L_2(\rho_d, D^d)$  space with finite

$$\|f\|_{L_2(\rho_d, D^d)}^2 = \int_{D^d} |f(\mathbf{x})|^2 \cdot \rho_d(\mathbf{x}) \, d\mathbf{x},$$

is a well defined Hilbert space. Here by  $\rho_d$  we mean

$$\rho_d(\mathbf{x}) = \prod_{j=1}^d \rho(x_j).$$

We then take

$$\mathcal{G}_d = L_2(\rho_d, D^d).$$

It is well known that the continuity of  $\mathcal{S}_1$  is equivalent to the continuity of the following integral operator

$$\mathcal{W}_1 := \mathcal{S}_1^* \circ \mathcal{S}_1 : H \rightarrow H, \quad \mathcal{W}_1(f)(x) = \int_d f(y) \cdot K(x, y) \cdot \rho(y) \, dy,$$

since then  $\|\mathcal{S}_1\|^2$  is equal to the largest eigenvalue of  $\mathcal{W}_1$ , i.e.,

$$C_0^2 = \max \{ \lambda : \lambda \in \text{spect}(\mathcal{W}_1) \}.$$

Then

$$1 \leq \|\mathcal{S}_d\|^2 \leq \left(1 + \frac{C_0^2}{d}\right)^d.$$

The condition (5) is now equivalent to

$$\int_D f(x) \cdot \rho(x) \, dx = 0 \quad \text{for all } f \in H,$$

which is satisfied by various spaces of periodic functions.

**2.3. Algorithms, Errors and Cost.** Since problems considered in this paper are defined over Hilbert spaces, we can restrict the attention to linear algorithms only, see e.g., [14], of the form

$$\mathcal{A}_n(f) = \sum_{j=1}^n L_j(f) \cdot a_j,$$

where  $L_j$  are continuous linear functionals and  $a_j \in \mathcal{G}_d$ . In the worst case setting considered in this paper, the error of an algorithm  $\mathcal{A}_n$  is defined by

$$\text{error}(\mathcal{A}_n; \mathcal{H}_d, \mathcal{G}_d) := \sup_{f \in \mathcal{H}_d} \frac{\|f - \mathcal{A}_n(f)\|_{\mathcal{G}_d}}{\|f\|_{\mathcal{H}_d}}.$$

So far, in the complexity study of problems with finitely many variables, it has been assumed that the cost of an algorithm is given by the number  $n$  of functional evaluations. We believe that, similar to problems with infinitely many variables, the cost of computing  $L(f)$  should depend on the *number of active variables* of  $L$ . More precisely, for given  $L \in \mathcal{H}_d^*$ , let  $h_L \in \mathcal{H}_d$  be its generator, i.e.,

$$L(f) = \langle f, h_L \rangle_{\mathcal{H}_d} \quad \text{for all } f \in \mathcal{H}_d.$$

Then  $h_L = \sum_{u \subseteq [1..d]} h_u$ ,

$$\text{Act}(L) := \left| \bigcup \left\{ \mathfrak{v} : h_{\mathfrak{v}} \neq 0, h_L = \sum_{u \subseteq [1..d]} h_u \right\} \right|$$

is the number of active variables in  $L$ , and the cost of evaluating  $L(f)$  is equal to

$$\$(\text{Act}(L)),$$

where  $\$ : \mathbb{N}_+ \rightarrow \mathbb{R}_+$  is a given *cost function*. The only assumptions that we make at this point are

$$\$(0) \geq 1 \quad \text{and} \quad \$(k) \leq \$(k+1) \quad \text{for all } k \in \mathbb{N}.$$

This includes

$$\$(k) = (k+1)^q, \quad \$(k) = e^{q \cdot k}, \quad \text{and} \quad \$(k) = e^{e^{q \cdot k}}$$

for some  $q \geq 0$ . Then the (*information*) *cost* of  $\mathcal{A}_n = \sum_{j=1}^n L_j(f) \cdot a_j$  is given by

$$\text{cost}(\mathcal{A}_n) := \sum_{j=1}^n \$(\text{Act}(L_j)).$$

The tractability results obtained so far for functions with finite numbers of variables correspond to  $\$ \equiv 1$ . In our opinion, it makes sense to assume that the cost function is at least linear, i.e.,

$$\$(k) \geq c \cdot (k+1), \quad k \in \mathbb{N}.$$

**2.4. Information Complexity and Tractability.** By *(information) complexity* we mean the minimal information cost among all algorithms with errors not exceeding a given error demand. That is, for  $\varepsilon \in (0, 1)$ ,

$$\text{comp}(\varepsilon; \mathcal{H}_d, \mathcal{G}_d) := \inf \{ \text{cost}(\mathcal{A}) : \text{error}(\mathcal{A}; \mathcal{H}_d, \mathcal{G}_d) \leq \varepsilon \}.$$

We now recall the definition of three kinds of tractabilities. For a detailed discussion of tractability concepts and results, we refer to excellent monographs [11, 12]. We stress however, that those results pertain to the constant cost function,  $\$ \equiv 1$ .

We say that the problem  $\mathcal{S}_d$  (or more precisely the sequence of problems  $\mathcal{S}_d$ ) is *polynomially tractable* if there exist  $c, p, q \geq 0$  such that

$$\text{comp}(\varepsilon; \mathcal{H}_d, \mathcal{G}_d) \leq c \cdot \frac{d^q}{\varepsilon^p} \quad \text{for all } \varepsilon \in (0, 1) \text{ and } d \in \mathbb{N}_+.$$

It is *strongly polynomially tractable* iff the above inequality holds with  $q = 0$ , and *weakly tractable* iff

$$\limsup_{d+1/\varepsilon \rightarrow \infty} \frac{\ln(\text{comp}(\varepsilon; \mathcal{H}_d, \mathcal{G}_d))}{d+1/\varepsilon} = 0.$$

When the problem is strongly polynomially tractable then

$$p^{\text{str}} := \inf \left\{ p : \sup_{\varepsilon, d} \varepsilon^p \cdot \text{comp}(\varepsilon; \mathcal{H}_d, \mathcal{G}_d) < \infty \right\}$$

is called the *exponent of strong tractability*.

There is also a concept of *quasi-polynomial tractability* introduced recently, see [3]. It is weaker than polynomial tractability and stronger than weak tractability. More precisely, the problem is quasi-polynomially tractable if there exist  $c, t \geq 0$  such that

$$\text{comp}(\varepsilon; \mathcal{H}_d, \mathcal{G}_d) \leq c \cdot \exp(t \cdot (1 + \ln(d)) \cdot (1 + \ln(1/\varepsilon))) \quad \text{for all } \varepsilon \in (0, 1) \text{ and } d \in \mathbb{N}_+.$$

This means that  $\text{comp}(\varepsilon; \mathcal{H}_d, \mathcal{G}_d) \leq c \cdot (e \cdot d)^{t \cdot (1 + \ln(1/\varepsilon))}$ . The significance of the quasi-polynomial tractability is that for some applications,  $d$  can be very large but  $\varepsilon$  need not be very small, say  $\varepsilon = 10^{-2}$ . Then the complexity of the problem is bounded by a polynomial in  $d$ .

As we shall prove in the next Sections, the problems considered in this paper are quasi-polynomially tractable even when the cost function  $\$$  is exponential in  $d$ .

### 3. RESULTS

**3.1. Number of Active Variables.** We are interested in a number  $m = m(\varepsilon, d)$  such that, for any  $f \in \mathcal{H}_d$ , the terms  $f_{\mathbf{u}}$  with  $|\mathbf{u}| > m$  can be neglected, i.e.,

$$(6) \quad \left\| \sum_{|\mathbf{u}| > m(\varepsilon, d)} f_{\mathbf{u}} \right\|_{\mathcal{G}_d} \leq \varepsilon \cdot \left\| \sum_{|\mathbf{u}| > m(\varepsilon, d)} f_{\mathbf{u}} \right\|_{\mathcal{H}_d}.$$

Hence, to approximate  $\mathcal{S}_d(f)$  with error bounded by  $\varepsilon\sqrt{2}$ , it is enough to use algorithms with functionals  $L_j$  that have  $\text{Act}(L_j) \leq m(\varepsilon, d)$ .

We first find  $m(\varepsilon, d)$  for the general tensor product space  $\mathcal{G}_d$  and next for the special case (1). To distinguish between the two cases, we will write respectively  $m_1 = m_1(\varepsilon, d)$  and  $m_2 = m_2(\varepsilon, d)$  instead of  $m = m(\varepsilon, d)$ .



3.1.1. *General Case.* For given  $\varepsilon \in (0, 1)$  and  $d \in \mathbb{N}_+$ , define

$$(7) \quad m_1 = m_1(\varepsilon, d) := \min \left\{ m : \sum_{k=m+1}^d \binom{d}{k} \cdot \left( \frac{C_0^2}{d} \right)^k \leq \varepsilon^2 \right\}.$$

Of course,  $m_1(\varepsilon, d)$  is well defined and is bounded by  $d$ .

**Proposition 1.** *For every  $d, \varepsilon \in (0, 1)$ , and  $f \in \mathcal{H}_d$ , (6) holds with  $m = m_1(\varepsilon, d)$  given by (7). Moreover,  $m_1(\varepsilon, d)$  is bounded from above by  $\min(d, M)$ , where  $M = M(\varepsilon)$  is the solution of*

$$\frac{(M+1)!}{C_0^{2 \cdot (M+1)}} = \frac{e^{C_0^2}}{\varepsilon^2}.$$

*In particular, there exists a constant  $C_1$  such that*

$$m_1(\varepsilon, d) \leq C_1 \cdot \frac{\ln(1/\varepsilon)}{\ln(\ln(1/\varepsilon))} \quad \text{for all } \varepsilon < e^{-e}.$$

*Proof.* Of course, (6) holds if  $m_1(\varepsilon, d) = d$ . Therefore we consider only the case when  $m_1 = m_1(\varepsilon, d) < d$ . We have

$$\begin{aligned} & \left\| \sum_{|u| > m_1} f_u \right\|_{\mathcal{G}_d} \leq \sum_{|u| > m_1} \|f_u\|_{\mathcal{G}_d} \leq \sum_{|u| > m_1} \|f_u\|_{H_u} \cdot C_0^{|u|} \\ & \leq \left[ \sum_{|u| > m_1} \gamma_{d,u}^{-1} \cdot \|f_u\|_{H_u}^2 \right]^{1/2} \cdot \left[ \sum_{|u| > m_1} \gamma_{d,u} \cdot C_0^{2 \cdot |u|} \right]^{1/2} \\ & = \left\| \sum_{|u| > m_1} f_u \right\|_{\mathcal{H}_d} \cdot \left[ \sum_{|u| > m_1} \gamma_{d,u} \cdot C_0^{2 \cdot |u|} \right]^{1/2} \\ & = \left\| \sum_{|u| > m_1} f_u \right\|_{\mathcal{H}_d} \cdot \left[ \sum_{k=m_1+1}^d \binom{d}{k} \cdot d^{-k} \cdot C_0^{2 \cdot k} \right]^{1/2} \\ & \leq \left\| \sum_{|u| > m_1} f_u \right\|_{\mathcal{H}_d} \cdot \varepsilon. \end{aligned}$$

This completes the proof of the first part. We now estimate the number  $m_1(\varepsilon, d)$ . Observe that, for any  $m < d$ , we have

$$\begin{aligned} & \sum_{k=m+1}^d \binom{d}{k} \cdot \left( \frac{C_0^2}{d} \right)^k = \sum_{k=m+1}^d C_0^{2 \cdot k} \cdot \frac{d \cdots (d-k+1)}{d^k \cdot k!} \\ & \leq \sum_{k=m+1}^d \frac{C_0^{2 \cdot k}}{k!} \leq \frac{C_0^{2 \cdot (m+1)}}{(m+1)!} \sum_{j=0}^{\infty} C_0^{2 \cdot j} \frac{(m+1)!}{(m+1+j)!} \\ & = \frac{C_0^{2 \cdot (m+1)}}{(m+1)!} \sum_{j=0}^{\infty} \frac{C_0^{2 \cdot j}}{j!} / \binom{m+1+j}{j} \leq \frac{C_0^{2 \cdot (m+1)} \cdot e^{C_0^2}}{(m+1)!}. \end{aligned}$$

This completes the proof.  $\square$

**Remark 1.** One can slightly improve the estimate of  $m_1(\varepsilon, d)$  by letting  $M = M(\varepsilon)$  to be the minimal integer such that  $C_0^2/(M+1) < 1$  and

$$(M+1)!/C_0^{2 \cdot (M+1)} \geq \frac{1}{\varepsilon^2 \cdot (1 - C_0^2/(M+1))}.$$

This is because the last sum in the proof above can be bounded as follows:

$$\sum_{j=0}^{\infty} \frac{C_0^{2 \cdot j}}{j!} / \binom{m+1+j}{j} \leq \sum_{j=0}^{\infty} \left( \frac{C_0^2}{m+1} \right)^j = \frac{1}{1 - C_0^2 \cdot (m+1)}.$$

We calculated the values of  $\lceil M(\varepsilon) \rceil$  for  $\varepsilon = 10^{-q}$  with  $q = 1, \dots, 10$  for the function approximation problem with the Wiener kernel on  $[0, 1]$  and  $\rho(x) \equiv 1$ . Recall that then  $C_0^2 = 1/2$ . These values are listed in the following table.

$q$	1	2	3	4	5	6	7	8	9	10
$\lceil M(10^{-q}) \rceil$	3	5	7	8	10	11	13	14	15	17

3.1.2. *Special Case (1).* We now investigate the number of active variables under the assumption (1). Then, for any  $k < d$ ,

$$\begin{aligned} \left\| \sum_{|u|>k} f_u \right\|_{\mathcal{G}_d}^2 &\leq C \cdot \sum_{|u|>k} \|f_u\|_{\mathcal{G}_d}^2 \\ &\leq C \cdot \sum_{|u|>k} C_0^{2 \cdot |u|} \cdot \gamma_{d,u}^{-1} \cdot \gamma_{d,u} \cdot \|f_u\|_{H_u}^2 \\ &\leq C \cdot \max_{\ell>k} (C_0^{2 \cdot \ell} \cdot d^{-\ell}) \cdot \left\| \sum_{|u|>k} f_u \right\|_{\mathcal{G}_d}^2. \end{aligned}$$

Therefore, for  $m_2 = m_2(\varepsilon, d)$  given by

$$(8) \quad m_2 := \begin{cases} 0 & \text{if } d < C_0^2 \text{ and } (C_0^2/d)^d \leq \varepsilon^2/C, \\ d & \text{if } d < C_0^2 \text{ and } (C_0^2/d)^d > \varepsilon^2/C, \\ \min(k : (C_0^2/d)^{k+1} \leq \varepsilon^2/C) & \text{otherwise,} \end{cases}$$

we have the following proposition.

**Proposition 2.** Suppose that (1) is satisfied. For every  $d, \varepsilon \in (0, 1)$ , and  $f \in \mathcal{H}_d$ , (6) holds with  $m = m_2(\varepsilon, d)$  given by (8). Moreover, for  $d \geq C_0^2$ ,

$$m_2(\varepsilon, d) \leq \min \left( d, \left\lceil \frac{\ln(C/\varepsilon^2)}{\ln(d/C_0^2)} \right\rceil - 1 \right) \quad \text{and} \quad m_2(\varepsilon, d) = O(\ln^{-1}(d)) \quad \text{as } d \rightarrow \infty.$$

**3.2. Changing Dimension Algorithm.** We consider in this section very special algorithms that are from the family of *changing dimension algorithms* introduced in [6] for integration and in [16, 17] for approximation of functions with infinitely many variables. As shown recently in [15], these algorithms yield polynomial tractability for weighted  $L_2$  approximation problems with infinitely many variables and general weights that have the decay greater than one.

These results are not applicable in this paper since the weights  $\gamma_{d,u} = d^{-|u|}$  have decay exactly one. However, these weights still allow for quasi-polynomial tractability and strong polynomial tractability if (1) holds.

More precisely, let  $\{(\lambda_{1,n}, \zeta_{1,n})\}_{n=1}^{\infty}$  be the set of eigenpairs of the operator

$$W_1 := S_1^* \circ S_1 : H \rightarrow H$$

for the class  $H$  of univariate functions. We assume that  $\lambda_{1,n}$  are monotonically decreasing to zero with a polynomial speed, i.e., that

$$(9) \quad \alpha := \text{decay}(\{\lambda_{1,n}\}_{n=1}^{\infty}) > 0.$$

Recall that the decay of a sequence of positive numbers  $a_n$  is defined by

$$\text{decay}(\{a_n\}_{n=1}^{\infty}) := \sup \left\{ t : \sum_{n=1}^{\infty} a_n^{1/t} < \infty \right\}.$$

For instance, the decay of  $a_n = \Theta(n^{-\beta} \cdot \ln^{\delta}(n))$  is equal to  $\beta$ . We also assume that  $\zeta_{1,n}$ 's form a complete orthonormal system in  $H$ . It is well known that the constant  $C_0$  is equal to the square-root of the largest eigenvalue of  $W_1$ , i.e.,

$$C_0 = \sqrt{\lambda_{1,1}}.$$

3.2.1. *General Case.* Consider the operator

$$W_u = S_u^* \circ S_u : H_u \rightarrow H_u$$

for the space  $H_u$ . Due to the tensor product structure of  $S_u$  and  $H_u$ , the eigenpairs of  $W_u$  are provided by the products of the eigenpairs for the univariate case. Let  $\{\lambda_{u,n}\}_{n=1}^{\infty}$  be the set of all the eigenvalues of  $W_u$  listed in the decreasing order,  $\lambda_{u,n} \geq \lambda_{u,n+1}$ . We now use a standard technique to estimate these eigenvalues. For that purpose note that, for any

$$\tau > 1/\alpha,$$

we have

$$\sum_{n=1}^{\infty} \lambda_{u,n}^{\tau} = [L(\tau)]^{|\mathbf{u}|} \quad \text{with} \quad L(\tau) := \sum_{n=1}^{\infty} \lambda_{1,n}^{\tau} < \infty.$$

Therefore the  $n$ th largest eigenvalue  $\lambda_{u,n}$  satisfies

$$n \cdot \lambda_{u,n}^{\tau} \leq [L(\tau)]^{|\mathbf{u}|}, \quad \text{i.e.,} \quad \lambda_{u,n} \leq \frac{[L(\tau)]^{|\mathbf{u}|/\tau}}{n^{1/\tau}}.$$

Let  $\zeta_{u,n}$  be the normalized eigenfunction corresponding to the eigenvalues  $\lambda_{u,n}$ . It is well known, see, e.g., [14], that the algorithm

$$A_{u,n}^*(f) := \sum_{j=1}^n \langle f, \zeta_{u,j} \rangle_{H_u} \cdot \zeta_{u,j}$$

have the minimal errors among all algorithms using  $n$  functional evaluations and

$$\text{error}(A_{u,n}^*; H_u, \mathcal{G}_u) = \sqrt{\lambda_{u,n+1}} \leq \left[ \frac{[L(\tau)]^{|\mathbf{u}|}}{n+1} \right]^{1/(2 \cdot \tau)}.$$

Since  $H_u$  are orthogonal subspaces of  $\mathcal{H}_d$ , the algorithms  $A_{u,n}^*$  are naturally extendable to  $\mathcal{H}_d$  and

$$A_{u,n}^* \left( \sum_{\mathbf{v} \subseteq [1..d]} f_{\mathbf{v}} \right) = A_{u,n}^*(f_u).$$

Moreover,

$$\text{cost}(A_{\mathbf{u},n}^*) \leq n \cdot \$(|\mathbf{u}|).$$

We are ready to define the algorithms  $\mathcal{A}_{\varepsilon,d}$  for the weighted space  $\mathcal{H}_d$ . For  $\varepsilon \in (0, 1)$ , let

$$(10) \quad \mathcal{A}_{\varepsilon,d}(f) := \langle f, 1 \rangle_{H_\emptyset} + \sum_{1 \leq |\mathbf{u}| \leq m_1(\varepsilon,d)} A_{\mathbf{u},n_{\mathbf{u}}}^*(f),$$

where

$$(11) \quad n_{\mathbf{u}} = n_{\mathbf{u},\varepsilon} := \left\lfloor \frac{[L(\tau)]^{|\mathbf{u}|}}{\varepsilon_u^{2 \cdot \tau}} \right\rfloor \quad \text{and} \quad \varepsilon_{\mathbf{u}} = \varepsilon_{\mathbf{u},d} := \frac{\varepsilon \cdot d^{|\mathbf{u}|/(2(1+\tau))}}{\sqrt{R}}$$

with

$$R = R(\varepsilon, d) := \sum_{k=1}^{m_1(\varepsilon,d)} \binom{d}{k} \cdot d^{-k \cdot \tau/(1+\tau)}.$$

Since  $n_{\mathbf{u}}$  depends on  $\mathbf{u}$  only via  $|\mathbf{u}|$ , we will sometimes write  $n_{|\mathbf{u}|}$  or  $n_\ell$  if  $|\mathbf{u}| = \ell$  instead of  $n_{\mathbf{u}}$ .

Note that

$$R \leq \sum_{k=1}^{m_1(\varepsilon,d)} \frac{d^k \cdot d^{-k \cdot \tau/(1+\tau)}}{k!} = \sum_{k=1}^{m_1(\varepsilon,d)} \frac{d^{k/(1+\tau)}}{k!} \leq m_1(\varepsilon, d) \cdot \frac{d^{\ell^*/(1+\tau)}}{(\ell^*)!},$$

where

$$\ell^* = \ell^*(\varepsilon, d) := \min \left( m_1(\varepsilon, d), \lfloor d^{1/(1+\tau)} \rfloor \right).$$

This follows from the fact that the sequence  $d^{k/(1+\tau)}/k!$  increases until  $k \leq d^{1/(1+\tau)}$ , and next starts to decrease, as can be easily verified. Hence

$$R^{1+\tau} \leq \begin{cases} d^{m_1(\varepsilon,d)/((m_1(\varepsilon,d)-1)!)^{1+\tau}} & \text{for } d > (m_1(\varepsilon,d))^{1+\tau}, \\ m_1(\varepsilon,d) \cdot e^{m_1(\varepsilon,d)} & \text{otherwise,} \end{cases}$$

where in the second case we replaced  $(\ell^*)!$  by  $(\ell^*/e)^{\ell^*}$  and used the fact that  $\ell^* \leq m_1(\varepsilon, d)$ . This means that

$$(R(\varepsilon, d))^{1+\tau} \leq C_1 \frac{\ln(1/\varepsilon)}{\ln(\ln(1/\varepsilon))} \cdot \varepsilon^{-C_1/\ln(\ln(1/\varepsilon))} \quad \text{if } m_1(\varepsilon, d) \geq d^{1/(1+\tau)}$$

and

$$(R(\varepsilon, d))^{1+\tau} \leq \varepsilon^{C_1/[(1+\tau) \cdot \ln(\ln(1/\varepsilon))]} \cdot d^{C_1 \cdot \ln(1/\varepsilon)/\ln(\ln(1/\varepsilon))} \quad \text{if } m_1(\varepsilon, d) < d.$$

Of course, in all the above estimates, we assume that  $\varepsilon < e^{-e}$ .

We now estimate the error of the algorithm  $\mathcal{A}_{\varepsilon,d}$ . Since  $\mathcal{A}_{\varepsilon,d}(f_{\mathbf{u}}) = 0$  for all  $f_{\mathbf{u}}$  with  $|\mathbf{u}| > m_1(\varepsilon, d)$ , we have

$$[\text{error}(\mathcal{A}_{\varepsilon,d}; \mathcal{H}_d, \mathcal{G}_d)]^2 = \sum_{1 \leq |\mathbf{u}| \leq m_1(\varepsilon,d)} \gamma_{d,\mathbf{u}} \cdot [\text{error}(A_{\mathbf{u},n_{\mathbf{u}}}^*; H_{\mathbf{u}}, \mathcal{G}_d)]^2 + \sum_{|\mathbf{u}| > m_1(\varepsilon,d)} \gamma_{d,\mathbf{u}} \cdot C_0^{2 \cdot |\mathbf{u}|}$$

The latter sum satisfies

$$\sum_{|\mathbf{u}| > m_1(\varepsilon,d)} \gamma_{d,\mathbf{u}} \cdot C_0^{2 \cdot |\mathbf{u}|} = \sum_{k=m_1(\varepsilon,d)+1}^d \binom{d}{k} \cdot \left( \frac{C_0^2}{d} \right)^k \leq \varepsilon^2,$$

whereas the former sum is bounded by

$$\begin{aligned}
\sum_{1 \leq |\mathbf{u}| \leq m_1(\varepsilon, d)} \gamma_{d, \mathbf{u}} \cdot \varepsilon_{\mathbf{u}}^2 &= \sum_{\ell=1}^{m_1(\varepsilon, d)} \binom{d}{\ell} \cdot d^{-\ell} \cdot \varepsilon_{\ell}^2 \\
&= \varepsilon^2 \cdot R^{-1} \cdot \sum_{\ell=1}^{m_1(\varepsilon, d)} \binom{d}{\ell} \cdot d^{-\ell} \cdot d^{\ell/(1+\tau)} \\
&= \varepsilon^2.
\end{aligned}$$

This means that

$$\text{error}(\mathcal{A}_{\varepsilon, d}; \mathcal{S}_d, \mathcal{H}_d) \leq \varepsilon \cdot \sqrt{2}.$$

We now estimate the cost of  $\mathcal{A}_{\varepsilon, d}$ :

$$\text{cost}(\mathcal{A}_{\varepsilon, d}) \leq \$ (0) + \sum_{1 \leq |\mathbf{u}| \leq m_1(\varepsilon, d)} \$ (|\mathbf{u}|) \cdot n_{\mathbf{u}} \leq \$ (0) + \$ (m_1(\varepsilon, d)) \sum_{1 \leq |\mathbf{u}| \leq m_1(\varepsilon, d)} n_{\mathbf{u}}$$

and

$$\begin{aligned}
\sum_{1 \leq |\mathbf{u}| \leq m_1(\varepsilon, d)} n_{\mathbf{u}} &= \sum_{\ell=1}^{m_1(\varepsilon, d)} \binom{d}{\ell} \cdot n_{\ell} \leq \varepsilon^{-2\tau} \cdot R^{\tau} \cdot \sum_{\ell=1}^{m_1(\varepsilon, d)} \binom{d}{\ell} \cdot \frac{[L(\tau)]^{\ell}}{d^{\ell \cdot \tau / (1+\tau)}} \\
&\leq \max(L(\tau), [L(\tau)]^{m_1(\varepsilon, d)}) \cdot \frac{R^{1+\tau}}{\varepsilon^{2\tau}}.
\end{aligned}$$

We summarize this in the following theorem.

**Theorem 1.** *Suppose that (9) holds. The approximation problem is quasi-polynomially tractable even if  $\$$  is an exponential function of  $d$ ,  $\$(d) = O(e^{q \cdot d})$ , and is weakly tractable even if  $\$(d) = O(e^{e^{q \cdot d}})$  for some  $q \geq 0$ . Moreover, for any  $\tau > 1/\alpha$ , the algorithms  $\mathcal{A}_{\varepsilon, d}$  have errors bounded by  $\varepsilon \cdot \sqrt{2}$  and cost bounded by*

$$\text{cost}(\mathcal{A}_{\varepsilon, d}) \leq \$ (0) + \$ (m_1(\varepsilon, d)) \cdot \max(L(\tau), [L(\tau)]^{m_1(\varepsilon, d)}) \cdot \frac{[R(\varepsilon, d)]^{1+\tau}}{\varepsilon^{2\tau}},$$

where  $m_1(\varepsilon, d)$  is given by (7), e.g.,

$$m_1(\varepsilon, d) \leq C_1 \cdot \frac{\ln(1/\varepsilon)}{\ln(\ln(1/\varepsilon))} \quad \text{for } \varepsilon < e^{-e},$$

and

$$[R(\varepsilon, d)]^{1+\tau} \leq \begin{cases} C_1 \frac{\ln(1/\varepsilon)}{\ln(\ln(1/\varepsilon))} \cdot \varepsilon^{-C_1 / \ln(\ln(1/\varepsilon))} & \text{if } m_1(\varepsilon, d) \geq d^{1/(1+\tau)}, \\ \varepsilon^{C_1 / [(1+\tau) \cdot \ln(\ln(1/\varepsilon))]} \cdot d^{C_1 \cdot \ln(1/\varepsilon) / \ln(\ln(1/\varepsilon))} & \text{otherwise.} \end{cases}$$

We believe that the result on quasi-polynomial tractability is sharp in general, i.e., there exist  $H$  and  $G_1$  such that the corresponding multivariate problem with weights  $\gamma_{d, \mathbf{u}} = d^{-|\mathbf{u}|}$  is only quasi-polynomially tractable. However, as we prove in the next section, Theorem 1 is not sharp when (1) holds.

3.2.2. *Special Case (1)*. We begin this section by assuming for a moment that

$$(12) \quad \left\| \sum_{\mathbf{u} \subseteq [1..d]} f_{\mathbf{u}} \right\|_{\mathcal{G}_d}^2 = \sum_{\mathbf{u} \subseteq [1..d]} \|f_{\mathbf{u}}\|_{\mathcal{G}_d}^2 \quad \text{for every } f \in \mathcal{H}_d,$$

which is a stronger assumption than (1). Similar spaces with norms satisfying (12) have been considered in [16, 17] for functions with infinitely many variables ( $d = \infty$ ) and some of the results below follow from [16].

As shown in [16], (12) allows for a simple characterization of the spectrum of

$$\mathcal{W}_d = \mathcal{S}_d^* \circ \mathcal{S}_d : \mathcal{H}_d \rightarrow \mathcal{H}_d$$

in terms of the spectrum of  $\mathcal{W}_1$ . Indeed, the eigenvalues of  $\mathcal{W}_d$  are given by

$$\gamma_{\mathbf{u}} \cdot \prod_{j \in \mathbf{u}} \lambda_{1,k_j}$$

for all  $\mathbf{u}$  and  $k_j \in \mathbb{N}$ . For  $\mathbf{u} = \emptyset$ , 1 is the corresponding eigenvalue.

Let  $\lambda_{d,n}$  ( $n \in \mathbb{N}_+$ ) be the eigenvalues of  $\mathcal{W}_d$  ordered so that

$$\lambda_{d,n} \geq \lambda_{d,n+1} \quad \text{for all } n.$$

Let  $\eta_{d,n}$  be the corresponding eigenfunctions that form a complete orthonormal system in  $\mathcal{H}_d$ . They also have a tensor product form and  $\eta_{d,n}$  corresponding to the eigenvalue  $\gamma_{\mathbf{u}} \prod_{j \in \mathbf{u}} \lambda_{1,k_j}$  has all the active variables listed in  $\mathbf{u}$ .

Define

$$\mathcal{A}_{\varepsilon,d}^*(f) := \sum_{j=1}^{n(\varepsilon,d)} \langle f, \eta_{d,j} \rangle_{\mathcal{H}_d} \cdot \eta_{d,j} \quad \text{with } n(\varepsilon,d) := \min \{k : \lambda_{d,k+1} \leq \varepsilon^2\}.$$

It follows from [16] that  $\mathcal{A}_{\varepsilon,d}^*$  is optimal for any cost function  $\$$ , i.e.,  $\text{error}(\mathcal{A}_{\varepsilon,d}^*; \mathcal{H}_d, \mathcal{H}_d) \leq \varepsilon$  and

$$\text{cost}(\mathcal{A}_{\varepsilon,d}^*) = \min \{ \text{cost}(\mathcal{A}) : \text{error}(\mathcal{A}; \mathcal{H}_d, \mathcal{G}_d) \leq \varepsilon \} = \text{comp}(\varepsilon; \mathcal{H}_d, \mathcal{G}_d).$$

Now  $\lambda_{1,1} = C_0^2$ ,

$$m_2(\varepsilon, d) = \min \left( d, \left\lceil \frac{\ln(1/\varepsilon^2)}{\ln(d/\lambda_{1,1})} \right\rceil - 1 \right),$$

and the functional evaluations  $\langle f, \eta_{d,j} \rangle_{\mathcal{H}_d}$  used by the algorithm have at most  $m_2(\varepsilon, d)$  active variables.

Note that, for every  $\delta > 0$ , we can bound  $m_2(\varepsilon, d)$  by

$$(13) \quad m_2(\varepsilon, d) \leq \max \left( \lambda_{1,1} \cdot e^{1/\delta}, \delta \cdot \ln(1/\varepsilon^2) \right).$$

Indeed, (13) trivially holds if  $d \leq \lambda_{1,1} \cdot e^{1/\delta}$ , and

$$\frac{\ln(1/\varepsilon)}{\ln(d/\lambda_{1,1})} < \frac{\ln(1/\varepsilon)}{\ln(e^{1/\delta})} = \delta \cdot \ln(1/\varepsilon^2) \quad \text{if } d > \lambda_{1,1} \cdot e^{1/\delta}.$$

Take now

$$\tau > 1/\alpha,$$

where, as before,  $\alpha = \text{decay}(\{\lambda_{1,n}\}_{n=1}^\infty) > 0$ . Using a standard technique, we get

$$\begin{aligned} \lambda_{d,k}^\tau \cdot k &\leq \sum_{j=1}^\infty \lambda_{d,j}^\tau = \sum_{u \subseteq [1..d]} \gamma_{d,u}^\tau \sum_{\mathbf{k} \in \mathbb{N}_+^{|u|}} \prod_{\ell=1}^{|u|} \lambda_{1,k_\ell}^\tau \\ &= \sum_{\ell=0}^d \binom{d}{\ell} \cdot \left( \frac{L(\tau)}{d^\tau} \right)^\ell = \left( 1 + \frac{L(\tau)}{d^\tau} \right)^d \\ &\leq e^{L(\tau) \cdot d^{1-\tau}} < \infty. \end{aligned}$$

Hence

$$\lambda_{d,k} \leq e^{L(\tau) \cdot d^{1-\tau}/\tau} \cdot k^{-1/\tau} \quad \text{and} \quad n(\varepsilon, d) \leq \left\lceil e^{L(\tau) \cdot d^{1-\tau}} \cdot \varepsilon^{-2\tau} \right\rceil - 1.$$

Note that the term  $e^{L(\tau) \cdot d^{1-\tau}}$  is bounded from above by  $e^{L(\tau)}$  if  $\tau \geq 1$ , and converges to 1 with increasing  $d$  if  $\tau > 1$ .

We return now to the original assumption (1). By replacing  $\varepsilon$  by  $\varepsilon/\sqrt{C}$  in all the formulas above, we get that  $\mathcal{A}_{\varepsilon/\sqrt{C},d}^*$  has the error bounded by  $\varepsilon$  when the norm in  $\mathcal{G}_d$  satisfies

$$\left\| \sum_{u \subseteq [1..d]} f_u \right\|_{\mathcal{G}_d}^2 = C \cdot \left\| \sum_{u \subseteq [1..d]} f_u \right\|_{\mathcal{H}_d}^2 \quad \text{for all } f \in \mathcal{H}_d.$$

Moreover, all upper bounds on the cost and errors provide corresponding upper bounds for norms that satisfy only (1), i.e., when the above equality is replaced by inequality.

This yields the following theorem.

**Theorem 2.** *Suppose that (1) and (9) hold. Then for any  $\tau > 1/\alpha$ ,*

$$\text{comp}(\varepsilon; \mathcal{S}_d, \mathcal{H}_d) \leq \$ \left( m_2(\varepsilon/\sqrt{C}, d) \right) \cdot \frac{e^{L(\tau) \cdot d^{1-\tau}}}{(\varepsilon/\sqrt{C})^{2\tau}}$$

with

$$m_2(\varepsilon/\sqrt{C}, d) \leq \min \left( d, \frac{\ln(C/\varepsilon^2)}{\ln(d/\lambda_{1,1})} \right).$$

Due to (13), the approximation problem is strongly polynomially tractable with the exponent

$$p^{\text{str}} \leq 2 \cdot \max(1, 1/\alpha).$$

even if  $\$(d) = O(e^{q \cdot d})$ , and is weakly tractable even if  $\$(d) = O(e^{e^{q \cdot d}})$  for some  $q \geq 0$ .

We now show that the upper bound on  $p^{\text{str}}$  is sharp if (12) holds.

**Proposition 3.** *If (12) and (9) hold then*

$$p^{\text{str}} = 2 \cdot \max(1, 1/\alpha).$$

*Proof.* Even for  $d = 1$ , we have  $\text{comp}(\varepsilon; H, \mathcal{G}_1) = \Omega(\varepsilon^{-2/\alpha})$ . Hence we only need to consider the case  $\alpha > 1$ . Suppose by the contrary that  $p^{\text{str}} = p$  for  $p < 2$ . The complexity of the problem with any cost function  $\$$  satisfying our assumptions is bounded from below by the complexity when  $\$(d) = 1$  for all  $d$ , and the latter complexity is fully determined by the eigenvalues of  $\mathcal{W}_d$ . That is, we have

$$\text{comp}(\varepsilon; \mathcal{H}_d, \mathcal{G}_d) \geq \$(0) \cdot \min \{ k : \lambda_{d,k+1} \leq \varepsilon^2 \}$$

for any cost function  $\$$ . Take any  $\widehat{p} \in (p, 2)$ . Then there is  $c(\widehat{p})$  such that

$$\lambda_{d,k} \leq c(\widehat{p}) \cdot k^{-2/\widehat{p}} \quad \text{for all } \varepsilon < 1 \text{ and } d \geq 1.$$

However, then, for any  $q > \widehat{p}$ ,

$$(14) \quad \sum_{k=1}^{\infty} \lambda_{d,k}^{q/2} \leq (c(\widehat{p}))^{q/2} \cdot \sum_{k=1}^{\infty} k^{-q/\widehat{p}} < \infty \quad \text{for all } d \geq 1.$$

Take  $q \in (\widehat{p}, 2)$ . As already explained,

$$\sum_{k=1}^{\infty} \lambda_{d,k}^{q/2} = \left(1 + \frac{L(q/2)}{d^{q/2}}\right)^d,$$

which converges to infinity as  $d \rightarrow \infty$ . This contradicts (14) and completes the proof.  $\square$

We apply Theorem 2 to the following  $L_2$  approximation problem.

**Example.** Consider  $K(x, y) = \min(x, y)$ ,  $D = [0, 1]$ , and  $\rho \equiv 1$ . It is well known that for the corresponding  $L_2$  approximation problem, we have  $\alpha = 2$ . Hence, for  $\$(d)$  at most exponential in  $d$ , we have strong tractability with the exponent

$$p^{\text{str}} \leq 2.$$

#### ACKNOWLEDGMENTS

The research presented in the paper was initiated during our stay at the *Hausdorff Research Institute for Mathematics*, University of Bonn, Summer 2011. We would like to thank the Institute and the organizers of *Analysis and Numerics of High Dimensional Problems* Trimester Program for their hospitality.

#### REFERENCES

- [1] J. Creutzig, S. Dereich, T. Müller-Gronbach, and K. Ritter, Infinite-dimensional quadrature and approximation of distributions, EMS Tracts in Mathematics Vol. **12**, Zürich, 2010.
- [2] M. Gnewuch, Infinite-dimensional integration on weighted Hilbert spaces, submitted.
- [3] M. Gnewuch and H. Woźniakowski, Quasi-polynomial tractability, *J. Complexity* **27** (2011), 312-330.
- [4] M. Hegland and V. Pestov, Additive models in high dimension, *ANZIAM J.* **46** (2005), 1205-1221.
- [5] F. J. Hickernell, T. Müller-Gronbach, B. Niu, K. Ritter, Multi-level Monte Carlo algorithms for infinite-dimensional integration on  $\mathbb{R}^{\mathbb{N}}$ , *J. Complexity* **26** (2010), 229-254.
- [6] F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski and H. Woźniakowski, Liberating the dimension, *J. Complexity* **26** (2010), 422-454. DOI: 10.1016/j.jco.2009.12.003.
- [7] F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski and H. Woźniakowski, On decompositions of multivariate functions, *Mathematics of Computation* **79** (2010), 953-966. DOI: 0.1090/S0025-5718-09-02319-9
- [8] P. Lévy, *Problèmes Concrets D'Analyse, Fonctionnelle*, Gauthier-Villars, Paris 1951.
- [9] B. Niu and F. J. Hickernell, Monte Carlo simulation of stochastic integrals when the cost function evaluation is dimension dependent, *Monte Carlo and Quasi-Monte Carlo Methods 2008* (P.L'Ecuyer and A. B. Owen, eds.), pp. 545-560, Springer Verlag, Berlin 2010.
- [10] B. Niu and F. J. Hickernell, T. Müller-Gronbach, and K. Ritter, Deterministic multi-level algorithms for infinite-dimensional integration on  $\mathbb{R}^{\mathbb{N}}$ , *J. Complexity* **27** (2011), 331-351. DOI: 10.1016/j.jco.2010.08.001
- [11] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems, Volume I: Linear Information*, EMS Tracts in Mathematics Vol. **6**, Zürich, 2008.



- [12] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems, Volume II: Standard Information for Functionals*, EMS Tracts in Mathematics Vol. **12**, Zürich, 2010.
- [13] L. Plaskota and G. W. Wasilkowski, Tractability of infinite-dimensional integration in the worst case and randomized settings, *J. Complexity* **27** (2011), 505-518.
- [14] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information-Based Complexity*, Academic Press, New York, 1988.
- [15] G. W. Wasilkowski, Liberating the dimension for  $L_2$  approximation, *J. Complexity* to appear.
- [16] G. W. Wasilkowski and H. Woźniakowski, Liberating the dimension for function approximation, *J. Complexity* **27** (2011), 86-110. DOI: 10.10.16/jco.2010.08.004
- [17] G. W. Wasilkowski and H. Woźniakowski, Liberating the dimension for function approximation: standard information, *J. Complexity* **27** (2011), 417-440. DOI: 10.10.16/j.jco.2011.02.002

M. Hegland, Mathematical Sciences Institute, The Australian National University, Canberra ACT 0200 Australia, email: markus.hegland@anu.edu.au

G. W. Wasilkowski, Department of Computer Science, University of Kentucky, Lexington, KY 40506, USA, email: greg@cs.uky.edu